UNIÓN
REVISTA IBEROAMERICANA DE EDUCACIÓN MATEMÁTICA

Marzo de 2009, Número 17, páginas 52 – 61
ISSN: 1815-0640

La función lineal obstáculo didáctico para la enseñanza de la regresión lineal

Héctor Agnelli, Patricia Konic, Susana Peparelli, Nora Zón, Pablo Flores

Resumen

Los conceptos de función lineal y regresión lineal generan dificultades en el proceso de enseñanza y aprendizaje de las matemáticas, particularmente en la escolaridad obligatoria. Con este trabajo intentamos poner de manifiesto que dichas dificultades pueden convertirse en obstáculos cuando se realizan tratamientos de situaciones escolarizables, que no ponen especial atención a la distinción entre los mencionados conceptos en el proceso de selección y análisis de situaciones y en la conducción del proceso de instrucción correspondiente

Abstract

The linear function and linear regression concepts generate difficulties in the teaching and learning processes, particularly in the compulsory school. In this work we try to show that these difficulties could become obstacles when treatments of school situations are carried out without recognizing those concepts in the selection process, in the analysis of the situations and in the ensuing teaching process.

Resumo

Os conceitos de função linear e regressão linear geram dificuldades no processo do ensino e aprendizagem das matemáticas, particularmente na escolaridade obrigatória. Com este trabalho tentamos manifestar que ditas dificuldades podem converter-se em obstáculos quando se realizam tratamentos de situações escolarizáveis, que não colocam especial atenção à distinção entre os mencionados conceitos no processo de seleção e análises de situações e na condução do processo de instrução correspondente.

1. Introducción

En los cursos de matemática se utiliza la función lineal como de uno de los ejemplos básicos del concepto de función. Dada la expresión de la función, $f(x)=mx+b$, en general se pasa a asociar esta función con su representación gráfica: la recta. Nominados ahora b y m como la ordenada al origen y la pendiente, se pone de manifiesto que las rectas y por lo tanto las funciones lineales, quedan individualizadas a partir del conocimiento de estos dos valores. Un ejercicio típico consiste en hallar la recta que determinan dos puntos, dadas sus coordenadas.

De esta manera queda establecida la ecuación de la recta al “calcular” la pendiente y la ordenada al origen. El problema es de naturaleza determinística.

El modelo de regresión lineal aparece en un contexto distinto, se utiliza para modelar fenómenos aleatorios. Con el modelo se intenta describir utilizando el lenguaje matemático el comportamiento de ese fenómeno aleatorio. Para ello se recurre a conceptos o expresiones matemáticas que adquieren significados distintos cuando se usan en el campo de la estadística. Se tiene un conjunto de puntos que expresan los valores que asumen dos características reales medibles y se quiere encontrar la relación que las vincula. Si bien el procedimiento que se utiliza para ello es de naturaleza determinística su interpretación a priori es de naturaleza aleatoria ya que distintos conjuntos de puntos, que representan a las mismas características, podrían originar distintas rectas. Cuestiones tales como parámetros, estimadores, variables aleatorias, independencia y distintas interpretaciones de la linealidad hacen su aparición en este tipo de problemas.

Consideramos que el uso de métodos determinísticos en la resolución de problemas matemáticos opera como un obstáculo para el abordaje de problemas de naturaleza aleatoria. Por lo que la noción de obstáculo se constituye en la herramienta didáctica que nos permite el estudio de esta problemática. Por ello sintetizaremos la posición que (Brosseau, 1983) sostiene al respecto.

Entre las características que Brosseau atribuye a lo que el llama un obstáculo destacamos la siguiente: un obstáculo es un conocimiento no una falta de conocimiento. Dicho conocimiento es adecuado para producir respuestas correctas en un determinado contexto; pero resulta insuficiente para abordar situaciones de otro contexto.

En sus investigaciones identifica tres tipos de obstáculos según su origen, Ontogenético, Epistemológico y Didáctico. El primero deriva de las características del desarrollo del niño, y que se ve reflejado en las concepciones que van desarrollando los alumnos durante su aprendizaje. El epistemológico está ligado a la naturaleza del conocimiento matemático y por último el obstáculo didáctico que surge de las elecciones didácticas con propósitos de enseñanza. En este trabajo nos centraremos en los obstáculos de origen didáctico, los cuales, como dijimos, proceden de “modos de enseñanza” y tienen un fuerte impacto en los aprendizajes de los alumnos. Son obstáculos que deben evitarse, lo que implica no solo identificarlos sino analizar las consecuencias que las distintas formas de enseñanza producen, Brousseau (1983).

Desde este marco consideramos que la comprensión de la modelización de un proceso estocástico en la enseñanza media, en particular el de la regresión lineal, debe ser abordado mostrando los límites de aplicación de la función lineal y de los procedimientos determinísticos, para el abordaje de los problemas de naturaleza aleatoria (Batanero y cls.1994). Si se introduce de manera precipitada la formalización de la regresión lineal para el estudio de fenómenos aleatorios bidimensionales, se puede generar un obstáculo didáctico al promover en el alumno una identificación de este modelo con el de la función lineal, que se aplica al estudio de fenómenos deterministas. Por ello, sostenemos que el primer paso para intentar evitar el obstáculo didáctico planteado es ***clarificar el uso operacional que tiene la función lineal en el marco de la estadística y poner en evidencia el rol que cumplen conceptos esencialmente determinísticos cuando son usados para la modelización de fenómenos aleatorios.***

Para desarrollar esta idea comenzaremos por analizar qué se entiende por función lineal en matemáticas y por modelo de regresión lineal en estadística, y cómo se relacionan ambos conceptos. Posteriormente mostraremos una tarea de regresión lineal que se propone en un libro de texto para la enseñanza polimodal argentina (estudiantes entre 14 y 15 años), en la cual pondremos en evidencia este obstáculo didáctico. Para salvarlo proponemos una enseñanza que tome en cuenta la naturaleza de ambos conceptos, por lo que recurrimos a las sugerencias que nos dan los estándares del NCTM (2000), para la enseñanza de la regresión lineal en las etapas 6-8 (niños de 12 a15 años) y 9-12 (16-18 años). Estas sugerencias, a nuestro juicio, proponen condiciones que permitirían evitar el obstáculo didáctico que estamos señalando. Finalmente se establecerán algunas conclusiones.

2. Marco referencial disciplinar

Buena porción de la actividad matemática puede ser descrita como proceso de modelización. Comenzaremos adoptando una concepción de modelo, siendo un punto de partida esencial para la discusión que pretendemos desarrollar. Según Henry (1997) "*Un modelo es una interpretación abstracta, simplificada e idealizada de un objeto del mundo real, de un sistema de relaciones o de un proceso evolutivo que surge de una descripción de la realidad*. Especialmente, "La modelización forma parte inseparable de la actividad estadística", Batanero (2001). Si bien en rigor no se puede pensar que un modelo sea real, en el sentido que sea una fiel representación de un fenómeno o sistema real, en estadística las inferencias están basadas en que el fenómeno se rige por un modelo determinado y la variación de datos se debe a otras causas, o a la misma naturaleza aleatoria del fenómeno.

Se emplean distintos tipos de modelos para estudiar situaciones deterministas de los que se emplean en fenómenos aleatorios. En las primeras los resultados están precisamente definidos, mientras que en los últimos los resultados muestran variabilidad debido a la presencia de factores desconocidos.

La modelización forma parte inseparable de la actividad estadística. En el proceso de modelización estadística se está interesado en descubrir patrones de comportamiento sistemático para datos obtenidos de experimentos u observaciones que contienen una componente aleatoria. Situaciones muy generales en las que se aplica dicho proceso están generadas a partir del interés que existe en explicar la variabilidad de una *característica* observada sobre unidades observacionales independientes y homogéneas para todos los aspectos de interés. Esta característica en el lenguaje de la estadística se llama *variable respuesta*. Se asume que existe algún mecanismo aleatorio desconocido que origina esa variabilidad y a partir de la información que entregan las observaciones realizadas, *los datos*, se construye un modelo simple que permita echar luz sobre esta variabilidad. En estas condiciones el modelo primario es una distribución de probabilidad. Pero en muchas situaciones, en las que no hay homogeneidad, se requiere algo más que una distribución de probabilidad para explicar la variación presente en los datos, la respuesta cambiaría además cuando cambia algunas de estas condiciones medibles, que operan sobre las unidades experimentales. Estas condiciones se llaman *variables explicativas* (explicativas de las respuestas). Así podría plantearse un modelo que tenga una parte sistemática además de una componente aleatoria. Uno de los modelos más simples es el llamado modelo de regresión lineal.

La idea central para la construcción del modelo de regresión lineal (Hamilton, 1992, Rawling, 1988), es que **la respuesta media de la variable Y** cambia con las condiciones de manera lineal, esto es de manera proporcional a la variación de X.

$$E(Y_i/x_i) = \alpha + \beta x_i \quad (1)$$

donde x_i es el valor que asume la variable explicativa *x* al medirla sobre el individuo *i.* Tanto α como β son desconocidos y se llaman los *parámetros* del modelo, ellos se conocerán o en el lenguaje estadístico se *estimarán* a partir de los datos.

El modelo (1) expresa como se espera que sea el comportamiento promedio de la variable respuesta *Y* al observarla repetidas veces bajo las mismas condiciones x_i. En rigor desde un punto de vista práctico, generalmente, se observa solo una vez y este caso se escribe

$$Y_i = \alpha + \beta x_i + \epsilon_i \quad (2)$$

Siendo ε un error de naturaleza aleatoria que origina la variación, en consecuencia también aleatoria, de los *Y* alrededor de su valor medio. Bajo estas condiciones los datos serán *n* pares de puntos $(y_i ; x_i)$ a partir de los cuales se debe determinar la recta (1).

Al observar la estructura de la expresión (2) surge que puede ser escrita como

$$Y_i = f(x_i) + \varepsilon_i$$

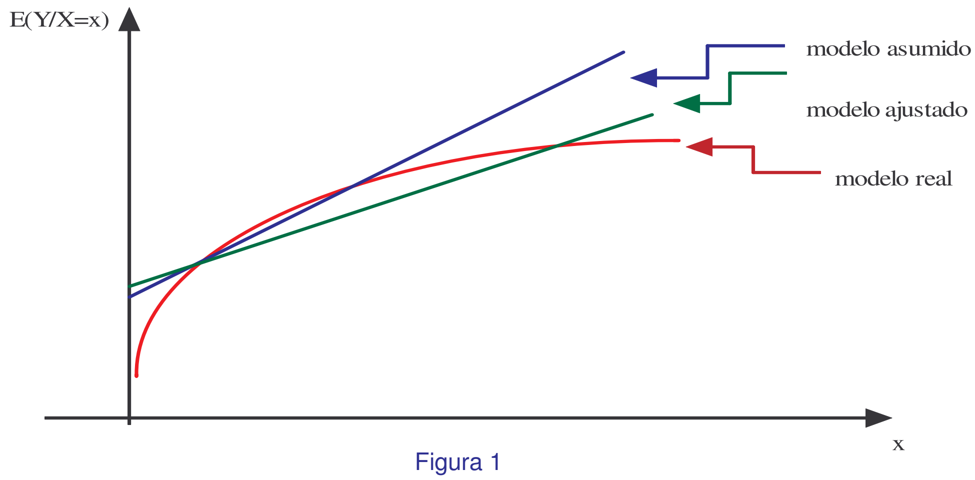

Figura 1

Donde ahora aparece la forma funcional habitual de la llamada función lineal pero como la parte sistemática del modelo

Cuando se trabaja con la modelización estadística y en particular con el modelo de regresión lineal están presentes tres modelos: el real y desconocido, el supuesto y el ajustado, como se muestra en la figura 1, de allí que en la estadística los parámetros del modelo resulten *estimados* por los valores que se *calculan* a partir de los datos que proporciona una muestra.

Por otra parte, cabe mencionar que, cuando en el contexto de la estadística se habla de un modelo lineal se hace referencia a la linealidad en los parámetros. A manera de ejemplos

$$E(Y_i/x_i) = \alpha + \beta x_i^2 \ \text{ y } \ E(\log Y_i/x_i) = \alpha + \beta \frac{1}{x_i} \text{ son lineales en } \alpha \text{ y } \beta$$

$$E(Y_i/x_i) = \alpha + \beta^2 x_i \ \text{ , no es lineal en } \beta.$$

De allí que algunos modelos lineales especifiquen que la relación gráfica entre la variable respuesta y la variable explicativa sea una recta, y otros expresen vínculos gráficos diferentes.

Con lo expuesto podemos establecer las siguientes diferencias entre la función lineal y la regresión lineal.

Función lineal	Regresión lineal
La variación de x e y es determinista	La variación de una o ambas, x e y es aleatoria
Relación funcional	Vínculo estadístico que extiende la noción de dependencia funcional.
El cálculo de los parámetros es independiente de los puntos seleccionados sobre la recta.	La estimación de los parámetros depende de los puntos seleccionados.

Dada una tabla de valores de *x* e *y* se obtiene una recta y dada esta es posible reproducir la tabla.	Dada una tabla de valores de *x* e *y* se obtiene una recta pero ella no permite reproducir la tabla.
El rol de variable independiente y dependiente es indistinto. $y = a_1x + b_1$ y $x = a_2y + b_2$ representan el mismo conjunto de puntos	Los roles de las variables explicativa y respuesta no son intercambiables. $y = \alpha_1 x + \beta_1$ y $x = \alpha_2 y + \beta_2$ no representan el mismo conjunto de puntos

3. Análisis de una situación

A los fines de poner de manifiesto el obstáculo didáctico que se genera al introducir de manera precipitada la regresión lineal para estudiar fenómenos aleatorios bidimensionales, dando lugar a que los alumnos lo identifiquen con el estudio de una función lineal, vamos a presentar y analizar una tarea de enseñanza extraída de un libro de texto de circulación actual (*Matemática 1- Ciencias, Primer año Polimodal,* 1999, p. 265). La que analizaremos en términos del uso que puede hacerse de la función lineal en un estudio de regresión lineal.

La tarea seleccionada se encuentra la final del capítulo destinado a estadística, en el texto mencionado, luego del desarrollo de una serie de nociones tales como: Muestras, técnicas de muestreo, distribuciones bidimensionales, recta de regresión, correlación y ecuación de la recta de regresión.

Tarea

La tabla recoge, de varias personas, el número de horas semanales que dedican a hacer deporte y el número de pulsaciones por minuto que tienen en reposo.

Hs. deporte	0	0	0	1	1	3	3	4	5	7
Pulsaciones	66	62	73	72	65	60	66	58	57	54

a) Dibuja la nube de puntos

b) ¿Hay dependencia funcional, estadística o casual?

c) La ecuación de la recta de regresión es: $y = -1{,}5x + 68$

- Estima el número de pulsaciones que tendrá una persona que dedica 2 horas semanales a hacer deporte.
- Estima el número de pulsaciones de un ciclista profesional que entrena 4 horas diarias. ¿Te parece razonable esta estimación?

Análisis de la tarea

Los autores del texto han elegido una situación aleatoria, próxima a los alumnos (deporte, colaborando en la educación para la salud). Los datos son números enteros, fáciles de manejar, de representar y corresponden a situaciones verosímiles. Las actividades que promueven son clásicas en el ámbito escolar, pero también en el estudio de regresión: dibujar nube de puntos, estudiar el tipo de dependencia y determinar el valor de una variable en función de la otra.

Pese a que la situación es interesante para la finalidad prevista, se ha planteado un texto conciso que hace que no se especifique qué rol cumplen cada una de las características que allí intervienen (horas semanales dedicadas al deporte y número de pulsaciones por minuto). Tampoco se explicita que se ha emprendido un estudio estadístico, ni el método de muestreo que se emplea.

En primer lugar se solicita dibujar la nube de puntos, para lo cual se provee una tabla del "tipo" de las que se presentan para graficar una función. Esta situación puede favorecer el obstáculo didáctico señalado, al relacionar de manera abusiva la regresión y el estudio de funciones.

En el inciso b) donde se presenta el interrogante crucial en esta temática, los datos aportados por la tabla "evidencian" no sólo una respuesta parcial, sino claramente de un tipo: no funcional. No se aborda, por lo tanto, el hecho que la vinculación estadística entre dos características es de naturaleza aleatoria y no surge de la no existencia de relación funcional. Con ello se contribuye a confundir dos aspectos relacionados con el término función: *relación unívoca* y *dependencia causal*. En el texto se pregunta por la segunda (dependencia funcional), pero se puede responder atendiendo a la otra consideración (no es unívoca). Esta confusión está muy relacionada con el obstáculo didáctico detectado, pues se utiliza el método de trabajo de las funciones lineales para responder a cuestiones de carácter estadístico.

En el inciso c) se presenta "la forma típica" de una función lineal sin identificar cuál es la característica denotada con x y cual la denotada con y, produciéndose así las siguientes contradicciones:

- No existe una relación funcional entre los datos y se presenta por medio de una función.
- No se puede reproducir la tabla a partir de la expresión analítica, procedimiento usual en la obtención de la gráfica de una función lineal.
- Bajo el supuesto de que el alumno infiera que la variable independiente x representa las horas de deporte y la dependiente y las pulsaciones, ¿qué interpretación le asignará a los valores de y que obtiene para cada uno de los valores de x de la tabla?

Por estas cuestiones es que la respuesta al pedido de "estimar" el número de pulsaciones será claramente de naturaleza determinística. Produciéndose en esta etapa un **afianzamiento** de la superposición conceptual función lineal-regresión lineal, calcular-estimar. Constituyéndose esto en una evidencia del tipo de obstáculo didáctico que caracterizamos.

El análisis realizado nos podría llevar a una postura pesimista respecto a la posibilidad de salvar el obstáculo comentado. Sin embargo hay propuestas didácticas que nos muestran que es posible atender a la diferencia conceptual de la regresión lineal respecto a la función lineal. En los estándares curriculares del NCTM (NCTM, 2000), hace algunas recomendaciones para el estándar que llama “Análisis de datos y probabilidad” que nos parecen interesantes. En primer lugar observamos que se diferencian las actividades a desarrollar en dos etapas: 6-8 (niños de 12-14 años, tercer ciclo de la Educación General Básica en Argentina), y 9-12 (15-18 años, educación polimodal en Argentina). En la primera el trabajo se realiza sobre la nube de puntos, sobre la que se dibuja una recta ajustada sin establecer criterios para ello, mientras que en la segunda etapa se llegan a probar con varias rectas y se ejercitan en la decisión sobre cuál es la que mejor se ajusta.

En concreto, para la etapa 6-8 se comienza por detectar y examinar dos características de una población, estudiando si existe alguna relación entre ellas, proporcionándoles datos a los alumnos y pidiendo que realicen una nube de puntos (tal como en la tarea analizada). Después, el profesor puede pedir que dibujen una recta que encaje aproximadamente en los datos, pero mostrando que no los cubre a todos. La determinación de la pendiente de esta recta (fase relacionada con el modelo que se emplea en el estudio de la función lineal), permite establecer la razón aproximada entre las variables estudiadas. (Utiliza el modelo determinístico haciendo énfasis siempre en que está estudiando de manera aproximada). También se propone emplear la nube de puntos para estudiar dos características en poblaciones diferentes. Como se observa, la nube de puntos es la referencia principal del estudio.

En la etapa 9-12, se centra en analizar las relaciones entre dos conjuntos de datos, incluyendo en este análisis el encontrar funciones que se ajusten bien a los datos. En sus sugerencias muestran que el proceso para llegar a establecer la regresión es el siguiente:

- Crear una nube de puntos
- Conjeturar qué modelo de función los aproxima mejor
- Buscar distintas restas de ajuste (en caso de que sea lineal)
- Determinar/decidir cuál es la que mejor se ajusta.

Los dos conjuntos de datos son, al comienzo y tal como ocurrió en la etapa 6-8, dos variables de la misma población, y posteriormente dos poblaciones.

Como podemos observar, los estándares muestran un cuidado especial en distinguir el razonamiento estadístico que se lleva a cabo en la regresión del formal que se lleva a cabo con funciones lineales.

4. Conclusiones

En los lineamientos curriculares de la escuela media, en distintos países, se incluyen contenidos de estadística desde edades tempranas, sin embargo consideramos que si no se toman recaudos, se corre el riesgo de subordinar el abordaje de problemas referidos a fenómenos aleatorios al trabajo con los conceptos matemáticos utilizados para el modelado estadístico. De esta manera la metodología esencialmente determinística de la matemática podría constituirse en un verdadero obstáculo para la comprensión de la naturaleza aleatoria de las situaciones que la estadística permite resolver.

Bibliografía

- Alvarez C.,. Alvarez F, Arribas A., Martinez S.y Ruiz A. (1999): *Matemática Ciencias, Primer año Polimodal*. Vicens Vives. Barcelona.
- Batanero C. (2001): "*Aleatoriedad, Modelización, Simulación".* Ponencia en X Jornadas sobre el Aprendizaje y la Enseñanza de las Matemáticas. Zaragoza.
- Batanero C.; Godino J. D., Green D. R., Colmes P.y. Vallecillos A (1994): *"Errores y dificultades en la comprensión de conceptos estadísticos elementales".* International Journal of Mathematics Education in Science and Technology, 25(4), 527-547. Recuperable en: [http://www.ugr.es/local/jgodino]
- Brousseau G. (1983): "*Les obstacles epistemologiques et les problèmes en mathématiques".* Rechearches en Didactique des mathématiques. 4(2), 164-198.
- Henry M. (1997): "*Notion de Modéle et modélization en l'enseignement". En Enseigner les probabilités au lycée* (pp. 77-84). Commision Inter-IREM, Reims
- Hamilton L. (1992): *Regression with Graphics*. Duxbury Press
- NCTM (2000): *Principios y Estándares para la Educación Matemática*. SAEM THALES, Sevilla.
- Rawling J. (1988): *Applied Regression Analysis*. Wadsworth & Brooks

Hector Agnelli, Licenciado en Matemáticas y Master en Estadística, profesor de Probabilidades y Estadísticas en la Universidad Nacional de Rio Cuarto. Argentina.

Patricia Konic, Profesora en Matemáticas, Master en Didáctica de la Matemática, docente de la Universidad Nacional de Río Cuarto, Argentina.
pkonic@exa.unrc.edu.ar

Susana Peparelli, Licenciada en Matemáticas, docente de la Universidad Nacional de Río Cuarto, Argentina.

Nora Zón, Profesora en Matemáticas, Master en Didáctica de la Matemática, docente de Enseñanza Media y de la Universidad Nacional de Río Cuarto. Argentina.

Pablo Flores, Licenciado en Matemáticas y en Ciencias de la Educación, Doctor en Matemáticas, profesor de Didáctica de la Matemática, Facultad de Ciencias de la Educación, Universidad de Granada, España.
pflores@ugr.es www.ugr.es/local/pflores